

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-172696
 (43)Date of publication of application : 23.06.2000

(51)Int.CI. G06F 17/30
 G06F 12/00

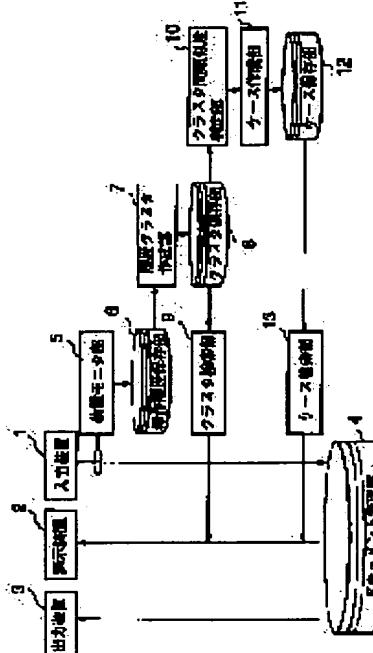
(21)Application number : 10-344183 (71)Applicant : TOSHIBA CORP
 (22)Date of filing : 03.12.1998 (72)Inventor : KIDOKORO KAZUAKI
 YODA NOBUHISA
 HARAGUCHI TATSUYA

(54) DOCUMENT MANAGING SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To extract the group of series of operation histories performed, based on the intention of a user out of the huge quantity of operation histories generated and stored at random with few processings.

SOLUTION: An operation monitor part 5 monitors user operation inputted from an input device 1, extracts the preparing/updating/reference processing of a document or document output processing, such as transmission/printing and records the document name of an operation object, time when the operation occurs, the identifier of a user who performs the operation and the user identifier at the transmission destination or the like in an operation history preserving part 6 as the operation history. Then, a history cluster preparing part 7 analyzes the operation histories stored in the operation history preserving part 6, for example divides it into small history sequences for each user time sequentially arranged within a predetermined timewise range, and records this divided operation history group in a cluster preserving part 8 as a cluster.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

This Page Blank (uspto)

[of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

This Page Blank (uspto)

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-172696

(P2000-172696A)

(43)公開日 平成12年6月23日 (2000.6.23)

(51)Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 06 F 17/30		G 06 F 15/401	3 1 0 D 5 B 0 7 5
12/00	5 3 1	12/00	5 3 1 J 5 B 0 8 2
		15/403	3 4 0 B

審査請求 未請求 請求項の数20 O.L (全 13 頁)

(21)出願番号 特願平10-344183

(22)出願日 平成10年12月3日 (1998.12.3)

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 城所 和明

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

(72)発明者 依田 信久

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

(74)代理人 100058479

弁理士 鈴江 武彦 (外6名)

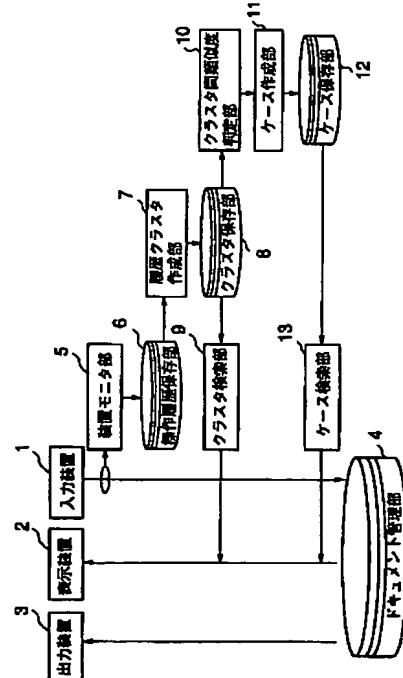
最終頁に続く

(54)【発明の名称】 ドキュメント管理システム

(57)【要約】

【課題】無作為に発生して蓄積される膨大な量の操作履歴の中から、ユーザの意図に基づいて行なわれた一連の操作履歴群を少ない処理量で抽出することができるドキュメント管理システムを提供する。

【解決手段】操作モニタ部1は、入力装置1から入力されるユーザ操作を監視して、ドキュメントの作成/更新/参照処理や、送信/印刷などのドキュメント出力処理を抽出し、操作対象のドキュメント名と、操作が発生した時刻、操作を行なったユーザ識別子および送信先のユーザ識別子などを操作履歴保存部6に操作履歴として記録する。そして、履歴クラスタ作成部7は、操作履歴保存部6に蓄積された操作履歴を解析し、たとえば予め定められた時間的な範囲内にあって時系列に並んだユーザ別の小さな履歴列に分割して、この分割後の操作履歴群をクラスタとしてクラスタ保存部8に記録する。



【特許請求の範囲】

【請求項1】 ドキュメントファイルに対する操作の履歴を蓄積する機能を有するドキュメント管理システムにおいて、前記蓄積された操作履歴を予め定められた規則により分類して前記操作履歴を要素とする集合体であるクラスタを作成するクラスタ作成手段を具備することを特徴とするドキュメント管理システム。

【請求項2】 前記クラスタ作成手段は、前記蓄積された操作履歴を操作者ごとに分けた後、この操作者ごとに分けられた操作履歴を所定の時間的範囲ごとに分けてクラスタを作成することを特徴とする請求項1記載のドキュメント管理システム。

【請求項3】 前記クラスタ作成手段は、前記蓄積された操作履歴を操作者ごとに分けた後、この操作者ごとに分けられた操作履歴を所定数ごとに分けてクラスタを作成することを特徴とする請求項1記載のドキュメント管理システム。

【請求項4】 前記クラスタ作成手段は、前記蓄積された操作履歴を操作者ごとに分けた後、この操作者ごとに分けられた操作履歴を特定の操作を区切りとして分けてクラスタを作成することを特徴とする請求項1記載のドキュメント管理システム。

【請求項5】 前記クラスタ作成手段が操作履歴を分類するための規則を設定するクラスタ設定手段をさらに具備することを特徴とする請求項1記載のドキュメント管理システム。

【請求項6】 前記クラスタ作成手段により作成された複数のクラスタの中から指定された検索条件を満たすクラスタを検出するクラスタ検索手段と、前記クラスタ検索手段により検出されたクラスタを表示するクラスタ表示手段とをさらに具備することを特徴とする請求項1、2、3、4または5記載のドキュメント管理システム。

【請求項7】 前記クラスタ作成手段により作成されたクラスタ間の類似度を算出する類似度算出手段をさらに具備することを特徴とする請求項1、2、3、4、5または6記載のドキュメント管理システム。

【請求項8】 前記類似度算出手段は、前記クラスタを構成する操作履歴の操作対象、操作者、操作日時および操作内容を含む属性情報からクラスタ間の類似度を算出することを特徴とする請求項7記載のドキュメント管理システム。

【請求項9】 前記類似度算出手段は、2つのクラスタに出現するドキュメントの一致数からいづれか一方のみに出現するドキュメントの不一致数を減算した値をクラスタ間の類似度とすることを特徴とする請求項7記載のドキュメント管理システム。

【請求項10】 前記類似度算出手段は、2つのクラスタから同一ドキュメントに対して行なわれた操作の組み合わせそれぞれの重みづけ値と、いづれか一方のクラ

タのみからいづれかのドキュメントに対して行なわれた操作の重みづけ値とをもち、

前記重みづけ値で重みづけた2つのクラスタに出現するドキュメントの一致数から前記重みづけ値で重みづけた前記いづれか一方のクラスタにのみ出現するドキュメントの不一致数を減算した値をクラスタ間の類似度とすることを特徴とする請求項7記載のドキュメント管理システム。

【請求項11】 前記類似度算出手段は、前記算出した値をさらにクラスタに含まれる操作履歴数で平均した値をクラスタ間の類似度とすることを特徴とする請求項9または10記載のドキュメント管理システム。

【請求項12】 前記類似度算出手段は、前記算出した値をさらにクラスタの発生時間の時間的距離に応じて重みづけた値をクラスタ間の類似度とすることを特徴とする請求項9または10記載のドキュメント管理システム。

【請求項13】 前記類似度算出手段が類似度を算出するために用いる重みづけ値を設定する重みづけ値設定手段をさらに具備することを特徴とする請求項10記載のドキュメント管理システム。

【請求項14】 前記類似度算出手段により算出されたクラスタ間の類似度を用いて、前記クラスタ作成手段が作成したクラスタを結合して一連の操作履歴群であるケースを作成するケース作成手段とをさらに具備することを特徴とする請求項7、8、9、10、11、12または13記載のドキュメント管理システム。

【請求項15】 前記ケース作成手段は、予め定められたしきい値を越える類似度をもった最も近接するクラスタ同士を結合させる手段を具備することを特徴とする請求項14記載のドキュメント管理システム。

【請求項16】 前記ケース作成手段は、单一のクラスタとすでに結合済みのクラスタ群に含まれるそれぞれのクラスタとの類似度の平均が予め定められたしきい値を越えるときに、前記单一のクラスタを前記クラスタ群に結合させる手段を具備することを特徴とする請求項14記載のドキュメント管理システム。

【請求項17】 前記ケース作成手段は、单一のクラスタとすでに結合済みのクラスタ群に含まれるそれぞれのクラスタとの類似度の平均を算出する際、クラスタ間の類似度にクラスタの発生時間の距離に応じた重みづけを行なうことを特徴とする請求項16記載のドキュメント管理システム。

【請求項18】 前記ケース作成手段は、单一のクラスタから予め定められた時間的範囲内に発生したクラスタのみを類似度判定の対象とすることを特徴とする請求項16または17記載のドキュメント管理システム。

【請求項19】 前記ケース作成手段がクラスタを結合するために用いるしきい値、重みづけ値または時間的範囲を設定するためのケース設定手段をさらに具備するこ

とを特徴とする請求項15、16、17または18記載のドキュメント管理システム。

【請求項20】前記ケース作成手段により作成された複数のケースの中から指定された検索条件を満たすケースを検出するケース検索手段と、前記ケース検索手段により検出されたケースを表示するケース表示手段とをさらに具備することを特徴とする請求項14、15、16、17、18または19記載のドキュメント管理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、複数のユーザから共有されるドキュメントに対する操作の履歴を蓄積する機能を有するドキュメント管理システムに係り、特に、無作為に発生して蓄積される膨大な量の操作履歴の中から、ユーザの意図に基づいて行なわれた一連の操作履歴群（業務シーケンス）を少ない処理量で抽出することができるドキュメント管理システムに関する。

【0002】

【従来の技術】大容量の記憶装置やネットワークの普及により、電子化されたドキュメントを用いた業務が一般化している昨今のオフィスにおいては、大量の電子ドキュメントの中から必要な情報をどれだけ効率的に取り出せるか、また、電子ドキュメントをどれだけ効果的に利用することができるかが、業務の効率に大きな影響を与える。

【0003】従来の木構造を用いたファイル管理方法では、オフィス業務においてドキュメントを管理するには機能が不十分であり、情報管理機能を補完し業務を効率化するために、たとえば検索機能を補完するための自然言語や類義語検索を用いた高機能なドキュメントの全文検索ツールや、予め定義されたドキュメントフローを用いて業務の効率化を図るため、すなわち、フロー管理機能を追加するためのワークフローシステムなどのドキュメント管理用アプリケーションプログラムが数多く開発されている。

【0004】これらのアプリケーションプログラムは、ドキュメントの検索や再利用を支援するツールとしてある程度の効果を挙げているが、ユーザのニーズを完全に満たすものではなく、これらのアプリケーションプログラムではカバーしきれない機能も必要とされている。たとえば、ドキュメントの全文検索アプリケーションプログラムでは、ユーザがドキュメントに含まれるキーワードを思い付かない場合には検索を行うことができない。また、ワークフローシステムでは、ドキュメントの利用方法が予め明確に定義されていなければならず、非定型のドキュメントが利用されることが多い日常のオフィス業務では使いにくいものであった。

【0005】また、最近では、ユーザのドキュメント操作履歴を用いてドキュメントの利用を支援しようとする

試みも見られる。この試みは、従来のドキュメント管理方法では利用できなかった「ドキュメントの利用履歴」という情報を用いてドキュメントを検索することを可能にしようとするものであり、たとえば特開平6-342451号公報に記載の文書管理装置では、記録されたユーザのドキュメント操作履歴を用いて、どのユーザがそのドキュメントをいつ頃参照したかなどの履歴情報を検索することによって、検索キーワードを思い付かなくても必要なドキュメントを発見することを可能としている。

【0006】

【発明が解決しようとする課題】ところで、前述した手法では、ドキュメントを検索するユーザは、そのドキュメントが誰に、どのように利用されたのかを推測して操作履歴を検索するが、ドキュメントを利用するユーザの立場から見ると、利用履歴に「Aさんはどのような順序でドキュメントにアクセスしたのか?」、「そのドキュメントを作成する時にどのような情報を参照したのか?」など、過去、ユーザが操作した時の操作意図などの情報が残されていた方が利用しやすいため、単に操作履歴を蓄積するだけでなく、操作を行なったユーザが自分の操作履歴を目的別に纏めたり、操作内容についてのコメントを残すことを可能にするシステムも提案されている。しかしながら、ユーザにコメントを残してもらう手法は、ユーザの負担が大きく、コメントの記述もが頻発するなどの問題があった。

【0007】一方、履歴を解析する手法としては、データマイニングの手法を応用してイベントログから頻繁に発生するシーケンシャルなパターンを発見する手法や、同じくデータマイニングの手法を応用してイベントログ間の相関関係を発見する手法などが存在するが、これらの手法は、イベントログの膨大な数の組み合わせのそれぞれが、イベントログ全体に対してどれくらいの頻度で発生するかを履歴データの解析手段として用いているために、パターン検出のための負荷が大きく、また、不連続なシーケンスを発見することができない、発生頻度の少ない相関関係が見つかりにくいなどの問題があり、複数のユーザが複数の意図で行なった操作履歴が入り交じった履歴データの解析が必要であるドキュメント管理システムでは適用が困難であった。

【0008】この発明は、このような実情に鑑みてなされたものであり、ドキュメントに対する操作の履歴の依存関係をユーザの負担なしに少ない処理量で抽出することができるドキュメント管理システムを提供することを目的とする。

【0009】

【課題を解決するための手段】前述した目的を達成するために、この発明においては、ドキュメントに対する操作の履歴をユーザ別に見た場合に、短い時間的範囲での操作は、特定の作業に関連した操作に限定されることに

注目し、連続した履歴データを短い時間的範囲でのユーザ別の連続した操作列（クラスタ）に区切り、さらに、この作成された操作列（クラスタ）間の類似度を判定して類似度の高い（同じ目的で行なわれた操作だと考えられる）操作列同士を結合し、より長い時間的範囲でのユーザ操作履歴列（ケース）を作成するようにしたものである。

【0010】この発明によれば、ユーザの負担なしに、かつ、データマイニングと比べて非常に少ない処理量で、膨大な量の操作履歴の中からユーザの意図に基づいて行なわれた一連の操作履歴群（業務シーケンスとなるクラスタおよびケース）を抽出することが可能となる。

【0011】

【発明の実施の形態】以下、図面を参照してこの発明の実施形態について説明する。

【0012】図1は、この実施形態に係る情報システムの機能ブロック図である。

【0013】入力装置1は、ユーザの操作をシステムに入力するためのキーボード、マウスまたはネットワークを介して接続される操作端末などである。表示装置2は、ユーザの操作に基づき、操作の結果やドキュメントを操作を行なったユーザに表示するディスプレイ装置などである。出力装置3は、ユーザの操作に基づき、ドキュメントをシステム外部に出力するためのプリンタ、ファックスまたはネットワーク装置などである。

【0014】ドキュメント管理部4は、記憶領域の管理や入出力装置、表示装置の制御を行なう通常のオペレーティングシステムであり、ネットワークを介して複数の入出力装置に接続され、複数のユーザに共有されて使用されている。このドキュメント管理部4は、入力装置1からのユーザの操作に基づき、ドキュメントの作成／編集／参照や、表示装置2や出力装置3への操作結果の表示／出力を行なう。また、このドキュメント管理部4では、ドキュメントのファイル名およびディレクトリ情報などの属性やユーザグループなどの情報が管理されている。さらに、このドキュメント管理部4は、ドキュメントに対するユーザの操作をモニタできるように拡張されている。

【0015】操作モニタ部5は、入力装置1から入力されるユーザ操作を監視して、ドキュメントの作成／更新／参照処理や、メール／ファックスによるドキュメントの送信またはドキュメントの印刷などのドキュメント出力処理を抽出し、操作対象のドキュメント名と、操作の発生した時刻、操作を行なったユーザ識別子および送信先のユーザ識別子などを後述する操作履歴保存部6に操作履歴として記録する。

【0016】操作履歴保存部6は、操作モニタ部5が検出した操作履歴を記録する。記録される操作履歴データの例を図2に示す。図2中、履歴ID（a1）は、操作履歴を管理するために割り当てられる番号である。操作

の種類（a2）は、記録された操作履歴の種類を表すものであり、「作成」、「更新」、「参照」、「削除」、「印刷」または「メール送信」など、ユーザが行なった操作の項目が発生した時間順に保存される。

【0017】ドキュメント名（a3）は、操作履歴とともに記録される操作対象のドキュメント名であり、この例では、オペレーティングシステムのファイルシステムでのディレクトリとファイル名とが記録されている。操作日時（a4）は、操作の発生した時刻が記録される。そして、ユーザ名（a5）は、操作履歴に残された操作を行なったユーザのユーザ識別子であり、ドキュメント管理部4で管理されているユーザ情報から取得されるものである。

【0018】また、履歴クラスタ作成部7は、操作履歴保存部6に蓄積された操作履歴を解析・分割し、時間順に並んだ小さな履歴列であるクラスタに変換する処理を行なう。クラスタは、ユーザが同じ目的で行なった処理の履歴が纏められることを意図している。この発明では、履歴データにはユーザがなぜその操作を行なったのかという情報は残されていないが、ユーザが複数の業務を並行して行なっていても、短い時間的範囲の操作に注目すると、特定の目的の操作を集中して行なうことが多いというドキュメント操作の特徴に着目し、複数のユーザの操作内容が記録された操作履歴列を操作を行なったユーザ別に分割し、さらに、ユーザごとの操作履歴を短い時間間隔内に発生した履歴列に分割する処理を行なう。

【0019】単に時間的範囲だけに注目して履歴を分割すると、別の意図で行なった操作履歴が同じクラスタに含まれてしまう確率が高いため（たとえば、「報告書を作成中に別件のメールを受信したため、文書作成業務を一時中断して返答メールを作成した。」のように異なる意図の操作が短い時間内に発生した場合など）、この履歴クラスタ作成部7では、経験的にユーザ操作の目的を変えるときに行なうことの多いと考えられる文書の保存操作や、メールの送信操作および文書の印刷操作などの特定の操作履歴を検出し、これらの操作が発生した時点でも操作履歴を分割する処理を行なう。

【0020】履歴クラスタ作成部7で作成されたクラスタは、クラスタ保存部8に記録される。図3は、作成されたクラスタの例を示すものである。図3中、クラスタID（b1）は、クラスタとして纏められた操作履歴群を管理するために割り当てられる番号である。たとえば、クラスタID121は、鈴木さんが10:30から10:43までに行なった操作を一つに纏めたものである。時間的に近接した操作であっても、別のユーザが行なった操作は別クラスタとして記録される（クラスタID122, 123）。なお、クラスタID124に含まれる鈴木さんの操作は、クラスタID121に含まれる操作と発生した時間が近接しているが、鈴木さんが履歴

ID330の操作で、操作の切れ目と考えられるドキュメントの印刷操作を行なっているため、クラスタ121とは分割されて別のクラスタとなっている。

【0021】クラスタ検索部9は、ユーザーの要求に応じてクラスタ保存部8に蓄積されたクラスタを検索・表示する処理を行なう。たとえば、ドキュメントA. docへのアクセスを含んだクラスタを検索すると、クラスタID121と122が見つかるが、クラスタにはドキュメントA. doc以外のドキュメントへのアクセスを含む複数の操作履歴が纏められているため、ドキュメントA. docとともにドキュメントC. pptが使われることがあるということがわかる。

【0022】クラスタ作成部7で作成されたクラスタは、時間的に近接したものを纏めただけであり、長い時間をかけて行なわれた業務や、業務の間に別の業務が発生したような場合は、履歴が途切れていることが考えられる（たとえば、前述した「報告書を作成中に別件のメールを受信したため、文書作成業務を一時中断して返答メールを作成した。」ような場合、メール受信前に行なわれていた文書作成業務の履歴と返答メール作成後に再開された文書作成業務の履歴とはユーザーの目的としては一つの纏まりである）。

【0023】クラスタ間類似度判定部10では、このようにクラスタ作成処理によって分断された操作履歴のクラスタ間の類似度を判定し、ケース作成部11によって類似度の高い（すなわち同じ目的で行われた可能性の高い）クラスタを連結して意味的につながりのあるユーザー操作列（ケース）を作成する。クラスタ間の類似度の判定には、「同じ業務を行なっている間には同じドキュメントが何度もアクセスされる」という経験則から、クラスタに同様のドキュメント操作が含まれているかどうかを用いる。また経験的に「同じ業務を行なっている最中には、同じドキュメントが何度も更新される」ことから、同じドキュメントの更新操作が含まれるクラスタは類似度を高くする、など操作の種類に応じて類似度に重み付けも行なう。

【0024】図4は、図3に示したクラスタ間の類似度を判定し、類似クラスタを連結して作成されたケースの例を示したものである。図4中、ケースID(c1)は、ケースとして連結されたクラスタ群を管理するために割り当てられる番号である。たとえば、ケースID32は、クラスタID121, 122, 124を纏めて発生時間順に並べ替えたものになっている。クラスタ121と122とには同じドキュメントA. docを参照した操作が含まれているため類似していると考えられる。この例では、複数のユーザーが共同で行なっている業務もケースとして纏められるように、別々のユーザーの行なったクラスタ間でも類似度の判定を行なっている（この場合は、鈴木さんと佐藤さんのクラスタが同じケースに纏められている）。クラスタ124の鈴木さんのクラスタ

は、クラスタ121とは共通のドキュメントが含まれていないが、クラスタ122で参照されているドキュメントC. pptへのアクセスが含まれているため、類似していると判断され同じケースに纏められている。作成されたケースは、ケース保存部12に蓄積される。

【0025】ケース検索部13は、クラスタ検索処理と同様に、ユーザーの要求に応じてケース保存部12に蓄積されたケースを検索・表示する処理を行なう。たとえばドキュメントA. docへのアクセスを含んだケースを検索すると、ケース32が発見される。ケースからは、クラスタの場合と同様に、ドキュメントA. docとともにドキュメントC. pptが使われることがあるということがわかるが、クラスタの場合と比べて複数の操作履歴が含まれるため、ドキュメントアクセスの発生回数からドキュメント間の関係に、より高い確信度があることがわかる。また、操作の発生パターンからドキュメントC. pptに関して鈴木さんが更新し佐藤さんが参照する関係にあることなど、クラスタでは明らかでなかった関係がわかる。

【0026】図5は、この実施形態に係る情報システムの構成を示す図である。

【0027】メインメモリ101は、実行中のプログラムとプログラムの実行に必要な制御用のデータとを記憶するためのものであり、ハードディスク102は、制御プログラムと、管理データおよび操作履歴とを記憶するものである。また、CPU103は、装置全体の動作を制御するものであり、システムが起動されるとハードディスク102に記憶されたプログラムをメインメモリ101上に読み出し、その内容にしたがって制御を行なう。

【0028】ユーザーとの対話処理には、入力装置としてキーボード104を使用し、入力された操作の結果をディスプレイ105に表示する。ドキュメントを出力する手段としては、プリンタ106が接続されている。また、この情報システムは、ネットワークインターフェース107を介してネットワークに接続され、ネットワークから入力された操作に対しての処理を行なったり、ネットワークに処理結果を出力すること、またはネットワークを介してドキュメントをメール/ファックスで送信することも可能である。

【0029】これらの各部は、システムバスにより接続されており、CPU103が動作制御するオペレーティングシステムにより制御される。そして、前述したこの実施形態の各機能は、同オペレーティングシステム上で動作するアプリケーションプログラムとして実現されている。

【0030】次に、この情報システムの処理内容を詳細に説明する。

【0031】図6は、この情報システムのメインフローである。

【0032】この情報システムは、その起動時、システム全体の初期化処理を行ない（ステップA1）、ユーザ操作の受け付けを開始する（ステップA2）。

【0033】ユーザがシステムに対してドキュメントの作成、参照、更新または出力などの操作要求を行なうと（ステップA3のYES）、この情報システムは、要求された操作内容を操作履歴保存部6に保存した上でユーザ操作の処理を行なう（ステップA4～ステップA5）。

【0034】ユーザからの要求が、処理パラメータの設定である場合（ステップA6のYES）、この情報システムは、図7に示すようなダイアログを表示して、処理パラメータの設定を行なう（ステップA7）。

【0035】図7中、クラスタ時間範囲（d1）は、一つのクラスタの時間的な大きさの最大値を設定するものであり、履歴データは、ここに設定された時間範囲を超えないクラスタに分割される。クラスタエントリ数（d2）は、一つのクラスタに含まれる操作履歴の数の最大値を定めるものであり、ドキュメントのアクセスがバースト的に発生したような場合でも、クラスタが大きくなりすぎるのを防ぐためのものである。

【0036】区切り操作（d3）は、操作の区切りとして検出すべき操作の種類を定めるものであり、この例では、ドキュメントの作成・更新操作とドキュメントの削除操作とがクラスタの区切りとして設定されている。

【0037】また、ユーザの要求がクラスタの検索処理であった場合（ステップA8）、この情報システムは、操作履歴保存部6に蓄積された履歴データがあれば、履歴データからクラスタを作成する履歴解析処理を行ない（ステップA9）、クラスタ保存部8に蓄積されたクラスタからユーザの要求にマッチするクラスタの検索と表示処理を行なう（ステップA10～ステップA11）。図8は、履歴データからクラスタを作成する処理の処理フローを示したものである。

【0038】クラスタ作成処理が開始されると、操作履歴保存部6に発生時間順に蓄積された履歴データを操作を行なったユーザ別に分割する（ステップB1）。次に、ユーザ別に分割された履歴列の中に、区切り操作として設定されている操作があれば、その操作の直後でさらに履歴列を分割する（ステップB2）。

【0039】また、分割された結果の履歴群のそれぞれについて、先頭の操作履歴の発生時間から最後の操作履歴の発生時間までが、登録された「クラスタ時間範囲」に収まるように、また、クラスタ内に含まれる操作履歴のエントリ数が、登録されたクラスタエントリ数以下になるように履歴群を分割する（ステップB3）。そして、ここまで処理で作成されたクラスタにクラスタIDをつけてクラスタ保存部8に記録し、操作履歴保存部6に蓄積された履歴データをすべて削除する（ステップB4）。

【0040】この実施形態では、一旦クラスタ作成のために処理された履歴データはクラスタ保存部8に移動されるため、検索時にクラスタ作成処理の対象となるのは前回のクラスタ作成処理以降に発生した履歴データのみであり、解析処理にかかる負荷を分散している。

【0041】図9は、ユーザがクラスタ検索要求を行なう際に検索パラメータを設定する画面の例である。この例では、操作を行なったユーザ、操作内容に含まれるドキュメント各、ドキュメントへの操作内容および操作の発生した時期（e1～e4）のうち、必要であるものを選択的に指定してクラスタの検索を行なう。指定できるパラメータの値が複数ある場合には、AND、ORの指定やワイルドカードを指定する。

【0042】図10は、ユーザが指定したパラメータを満たすクラスタの検索結果を表示する画面の例である。この例では、「鈴木さん」が、「更新または印刷」を行なった操作履歴を含み、「12/23～12/31」の間に発生したクラスタを一覧で表示している。

【0043】そして、ユーザからの要求がシステムの終了処理であった場合（ステップA12のYES）、この情報システムは、管理データおよび履歴データをハードディスク102上に退避して、終了処理を実行する（ステップA13）。

【0044】このように、この実施形態の情報システムによれば、操作履歴保存部6に蓄積された膨大な量の操作履歴を、少ない処理量で、ユーザが同じ目的で行なった処理の履歴が纏められたクラスタに変換することができ、ユーザは、このクラスタを効率的に検索し参照することが可能となる。

【0045】次に、前述したクラスタの作成に加えて、さらに類似クラスタを連結させてケースを作成する場合を説明する。この場合、前述の処理に加えて、クラスタ同士を比較してクラスタ間の類似度を判定する処理と、類似したクラスタを連結してケースを作成する処理とが追加される。

【0046】図11は、クラスタ間の類似度を判定する処理のフローを示している。

【0047】この場合、この情報システムは、まず、クラスタ保存部8に蓄積されたクラスタ間の類似度をすべて0に初期化する（ステップC1）。次に、クラスタ群の中から一つのクラスタを選択し（ステップC2）、そのクラスタと「最大ケース時間範囲」に設定された範囲内に発生したクラスタすべてのクラスタそれぞれの類似度を計算する（ステップC3）。

【0048】類似度の計算は、主にクラスタに含まれるドキュメントおよびドキュメントに対する操作の種類の比較によって行なう。図12は、クラスタ間の類似度の計算に用いるパラメータを設定する画面の例を示したものである。

【0049】図12中、ケース時間範囲（f1）は、前

述したように、ケースとして連結することが可能なクラスタの発生時間範囲を設定するものである。この時間を設定することで不必要に長いケースを作成することがなくなり、また、時間範囲内のクラスタだけの比較に限定することで比較処理の負荷を抑えることができる。

【0050】類似度判定パラメータ(f_2)は、クラスタ間の類似度を計算するために用いられる。たとえば、図12中に設定された値を用いると、二つのクラスタを比較して、あるドキュメントが双方のクラスタで参照されていた場合は類似度を5ポイント増加し、あるドキュメントが双方のクラスタで更新されていた場合は類似度を100ポイント増加する。また、あまり関係のないクラスタが類似度が高いと判定されるのを防ぐために、一方のクラスタで参照されているドキュメントがもう一方のクラスタではアクセスされていない場合に、逆に類似度を2ポイント減少する。

【0051】たくさんのドキュメントアクセス履歴を含むクラスタはそれだけ多くのクラスタと共通したアクセス履歴を含むことが多く、類似度が不必要に高くなってしまうことを避けるため、クラスタに含まれる共通履歴の類似度の累積として計算されたクラスタ間の類似度を最後に、そのクラスタに含まれる履歴の数で除算してクラスタ間の類似度とする。

【0052】また、図12で設定される残り2つのパラメータは、後のケース作成処理で用いるものである。クラスタ連結しきい値(f_3)は、クラスタ間の類似度との比較に用い、2つのクラスタを連結するかどうかの判断に用いる。複数ユーザのクラスタを連結するかどうかの設定(f_4)は、連結するクラスタを同一ユーザの行なった操作に限定するかどうかを制御するために用いる。

【0053】そして、この情報システムは、すべてのクラスタについて、前述した類似度計算を実行する(ステップC4)。

【0054】図13は、類似度をもとにクラスタを連結してケースを作成する処理のフローを示したものである。

【0055】まず、この情報システムは、クラスタ保存部8に蓄積されたクラスタから最も過去に発生した(クラスタの先頭の操作履歴の発生時間が最も古い)クラスタ(クラスタAとする)を選択し、新規に作成したケース(ケースAとする)にそのクラスタの内容を登録する(ステップD1)。

【0056】次に、クラスタ保存部8の残りのクラスタの内、ケースAの先頭の操作履歴の発生時間から、「ケース時間範囲」内に発生したクラスタを、発生時間の古い順に選び(クラスタBとする)、クラスタAまたはケースAと結合するかどうかの判定を行なう(ステップD2)。ケース作成処理のパラメータとして、「複数ユーザのクラスタを連結しない」と指定されている場合に

は、クラスタBとして選択するのはクラスタAの操作を行なったユーザと同じユーザが行ったクラスタに限定する。

【0057】判定方法としては、たとえば、クラスタAとクラスタBの類似度が、「クラスタ連結しきい値」よりも大きい場合クラスタが連結可能であると判定する。あるいは、クラスタBと、すでにケースに組み込まれているすべてのクラスタとの類似度を平均して、ケースAとクラスタBの類似度とし、「クラスタ連結しきい値」よりも大きい場合にクラスタが連結可能であると判定する。さらに、クラスタBと、すでにケースに組み込まれているすべてのクラスタとの類似度を平均する際、ケース内のクラスタの発生時間とクラスタBの発生時間の差を類似度の重みづけとして用いることも有用である。たとえば、クラスタ1とクラスタ2との発生時間の差を「ケース時間範囲」で割ったものをクラスタ1～2間の類似度に掛けたものをクラスタ1～2の重みづけされた類似度として用いる。そして、この重みづけされた類似度が「クラスタ連結しきい値」よりも大きい場合にクラスタが連結可能であると判定する。

【0058】クラスタBがケースAに連結可能であると判定されると(ステップD2のYES)、クラスタBをケースAに連結する(ステップD3)。また、クラスタBを新たにクラスタAとし、ステップD2に戻って次のクラスタの選択、結合可能性の判定を続ける。

【0059】一方、ケースAに連結可能なクラスタが見つからなかった場合(ステップD2のNO)、ケースAに含まれる操作履歴を時間順に並べ替え、ケースIDを付けてケース保存部12に登録し、ケースAに登録されたクラスタが他のケースに含まれないように登録済みのクラスタをクラスタ保存部8から削除する(ステップD4)。

【0060】そして、すべてのクラスタについてのケース化の判断が終了するまで、ステップD1～ステップD4の処理を繰り返す。

【0061】なお、ケース作成処理を行なった場合のユーザの履歴検索操作は、ケース保存部12に対して実行される。検索対象、表示内容がクラスタではなくケースであることを除けば、処理の内容は前述のクラスタ検索・表示処理と同様に行なわれる。

【0062】また、前述の方法では、ケース作成処理は、ユーザがケースの検索を要求するたびに行なわれるため、ケース検索処理が頻繁に行なわれるとケース作成処理がごく少数の履歴クラスタに対して行なわれることになり、十分連結されていない短いケースが作成されてしまうことになる。これを避けるためには、ケースを登録する際に、ケース作成時の時間とケースに登録された最初の操作履歴の発生時間との差が「ケース時間範囲」よりも小さいケースは、ケースとして登録せずにクラスタに戻してクラスタ保存部8に戻す処理を行なえばよ

い。

【0063】このように、この実施形態の情報システムによれば、クラスタ保存部8に蓄積されたクラスタを、互いに類似するクラスタ同士が連結されたより大きな集合体であるケースに変換することができ、ユーザは、このケースを効率的に検索し参照することが可能となる。

【0064】

【発明の効果】以上詳述したように、この発明によれば、ドキュメントファイルに対して行なわれる操作の特徴を用いて、履歴の蓄積・解析に関してユーザに負担をかけることなく、小さい解析負荷で意味のある履歴集合を作成することが可能となる。

【0065】また、この作成した履歴集合をユーザに表示することで、履歴の直接検索では得られないドキュメント、ユーザおよび操作内容の関連を検出することが可能となる。

【0066】さらに、クラスタを用いて類似度を判定することで、少ない負荷で類似度の判定を行ない、操作履歴のグループ化を行なうことが可能となる。

【図面の簡単な説明】

【図1】この発明の実施形態に係る情報システムの機能ブロック図。

【図2】同実施形態の操作履歴保存部に記録される操作履歴データを例示する図。

【図3】同実施形態の履歴クラスタ作成部により作成されるクラスタを例示する図。

【図4】同実施形態のケース作成部により作成されるケースを例示する図。

【図5】同実施形態の情報システムの構成を示す図。

【図6】同実施形態の情報システムのメインフロー。

【図7】同実施形態のクラスタ作成に用いられる処理パラメータの設定画面を例示する図。

【図8】同実施形態の履歴データからクラスタを作成する処理の処理フロー。

【図9】同実施形態のクラスタ検索要求を行なう際に検索パラメータを設定する画面を例示する図。

【図10】同実施形態の指定されたパラメータを満たすクラスタの検索結果を表示する画面を例示する図。

【図11】同実施形態のクラスタ間の類似度を判定する際の処理フロー。

【図12】同実施形態のクラスタ間の類似度の計算に用いるパラメータを設定する画面を例示する図。

【図13】同実施形態の類似度をもとにクラスタを連結してケースを作成する際の処理フロー。

【符号の説明】

1…入力装置

2…表示装置

3…出力装置

4…ドキュメント管理部

5…装置モニタ部

6…操作履歴保存部

7…履歴クラスタ作成部

8…クラスタ保存部

9…クラスタ検索部

10…クラスタ間類似度判定部

11…ケース作成部

12…ケース保存部

13…ケース検索部

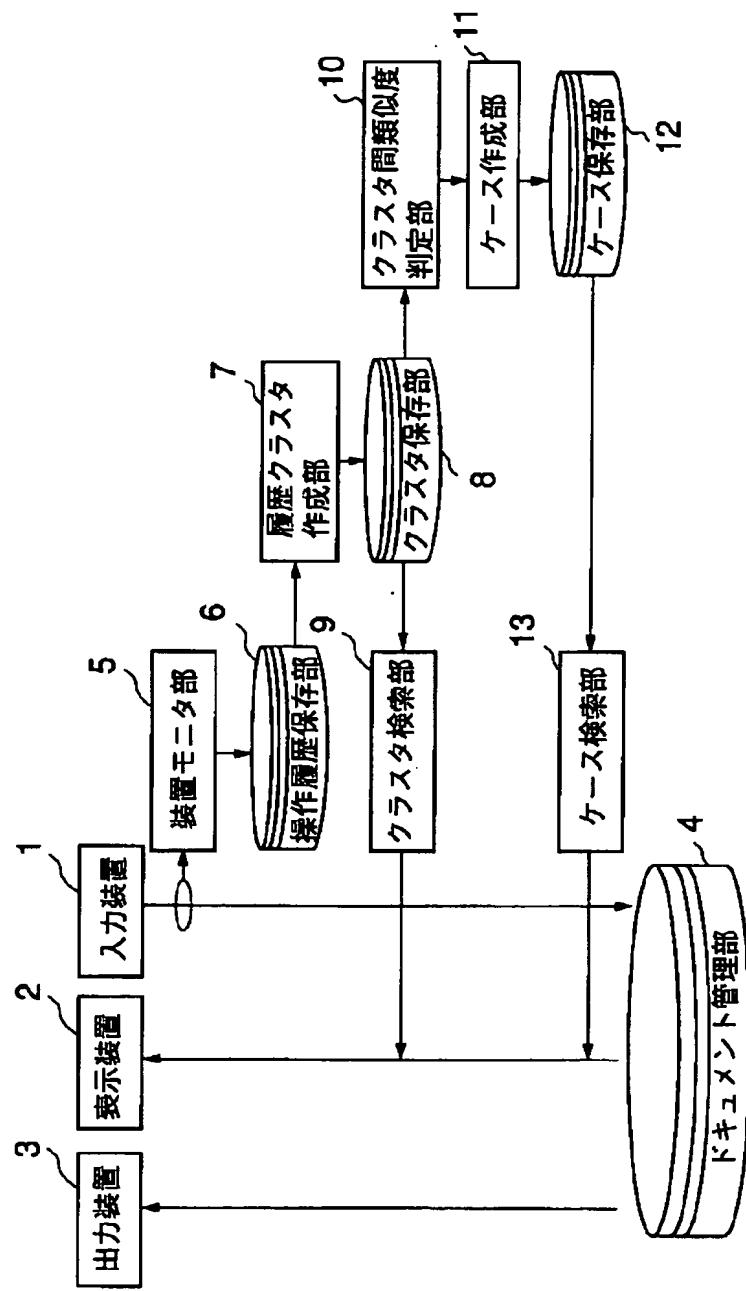
【図2】

履歴ID	操作の種類	ドキュメント名	操作日時	ユーザ名
1	1	1	1	1
324	参照	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:30:20	鈴木
325	参照	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:31:05	佐藤
326	参照	F:\drBV\ドキュメントB.txt	1997/12/24 10:32:20	高橋
327	参照	F:\drBV\ドキュメントC.ppt	1997/12/24 10:35:21	佐藤
328	更新	F:\drBV\ドキュメントD.doc	1997/12/24 10:41:12	高橋
329	更新	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:43:20	佐藤
330	印刷	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:43:48	鈴木
331	更新	F:\drBV\ドキュメントC.ppt	1997/12/24 10:50:45	鈴木
1	1	1	1	1

【図3】

クラスタID	履歴ID	操作の種類	ドキュメント名	操作日時	ユーザ名
121	1	1	1	1	1
	324	参照	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:30:20	鈴木
	330	印刷	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:43:45	鈴木
122	325	参照	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:31:05	佐藤
	327	参照	F:\drBV\ドキュメントC.ppt	1997/12/24 10:35:21	佐藤
	329	更新	C:\MaxDV\ドキュメントA.doc	1997/12/24 10:43:20	佐藤
123	326	参照	F:\drBV\ドキュメントB.txt	1997/12/24 10:32:20	高橋
	328	削除	F:\drBV\ドキュメントD.doc	1997/12/24 10:41:12	高橋
124	331	更新	F:\drBV\ドキュメントC.ppt	1997/12/24 10:50:45	鈴木
1	1	1	1	1	1

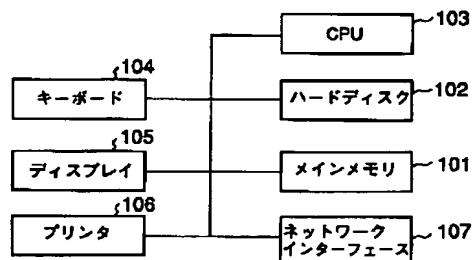
【図1】



【図4】

ケースID	クラスタID	履歴ID	操作の種類	ドキュメント名	操作日時	ユーザ名
32	121	324	参照	C:\text\ドキュメントA.doc	1997/12/24 10:30:20	鈴木
		325	参照	C:\text\ドキュメントA.doc	1997/12/24 10:31:05	佐藤
		327	参照	F:\dirB\ドキュメントC.ppt	1997/12/24 10:35:21	佐藤
	123	329	更新	C:\text\ドキュメントA.doc	1997/12/24 10:43:20	佐藤
		330	印刷	C:\text\ドキュメントA.doc	1997/12/24 10:43:45	鈴木
		331	更新	F:\dirB\ドキュメントC.ppt	1997/12/24 10:50:45	鈴木
33	123	326	参照	F:\dirA\ドキュメントB.txt	1997/12/24 10:32:20	高橋
		328	削除	F:\dirA\ドキュメントD.doc	1997/12/24 10:41:12	高橋

【図5】



【図7】

クラスタ時間範囲 最大 [500] 秒
 クラスタエントリ数 最大 [100]
 区切り操作
 ■ 作成・更新 ■ 出力 (印刷・FAX)
 ■ 削除 ■ メール送信
 [OK] [Cancel]

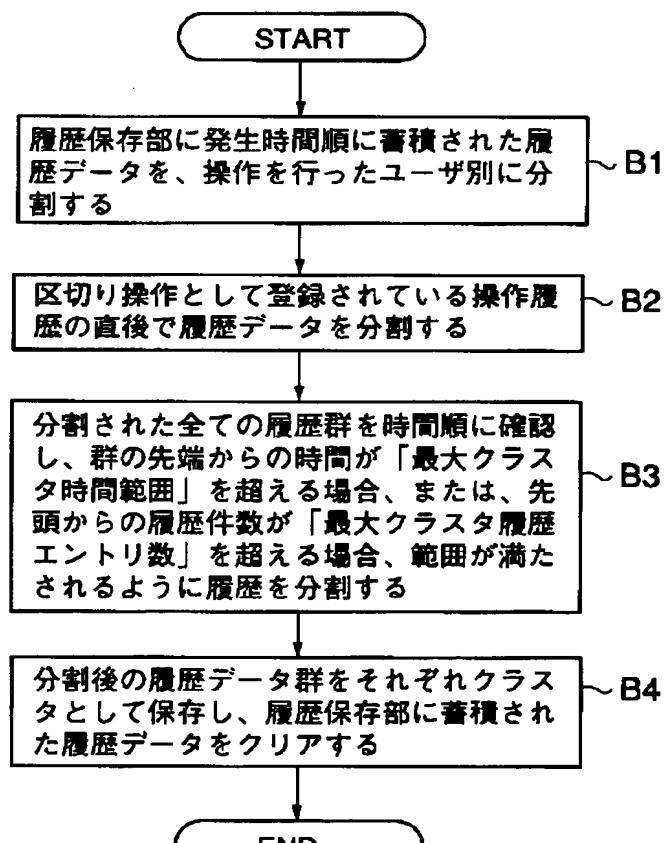
【図9】

クラスタ検索パラメータ
 ユーザ [鈴木] e1
 ドキュメント名 [] e2
 操作内容 [更新 or 印刷] e3
 発生時間 [1997/12/23~1998/12/31] e4
 [検索] [Cancel]

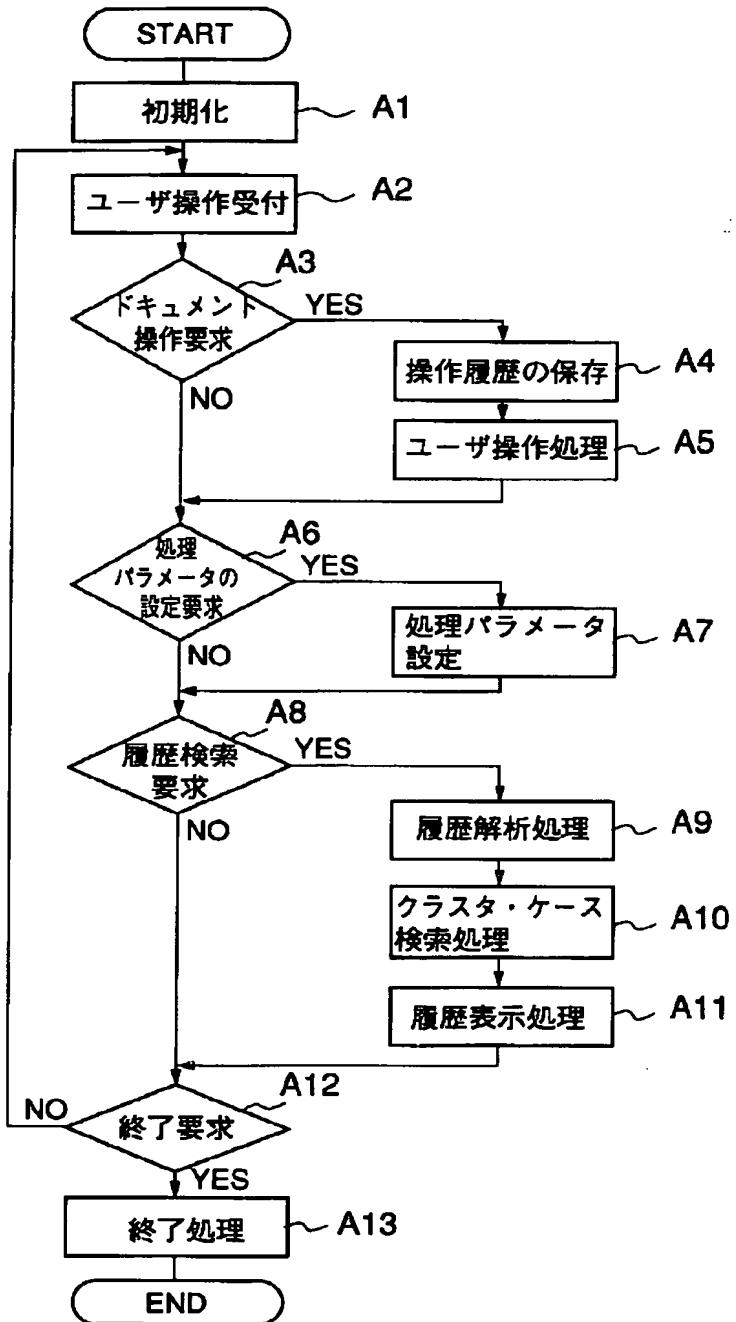
【図12】

ケース時間範囲 最大 [5000] 秒 f1
 類似度判定パラメータ f2
 同じドキュメントを参照 [5] f3
 同じドキュメントを作成・更新 [100] f4
 同じドキュメントを出力 [30] f5
 その他同一ドキュメントを操作 [10] f6
 異なるドキュメントの操作 [-2] f7
 クラスタ連結しきい値 [20] f8
 複数ユーザのクラスタを連結 [YES] f9
 [OK] [Cancel]

【図8】



【図6】



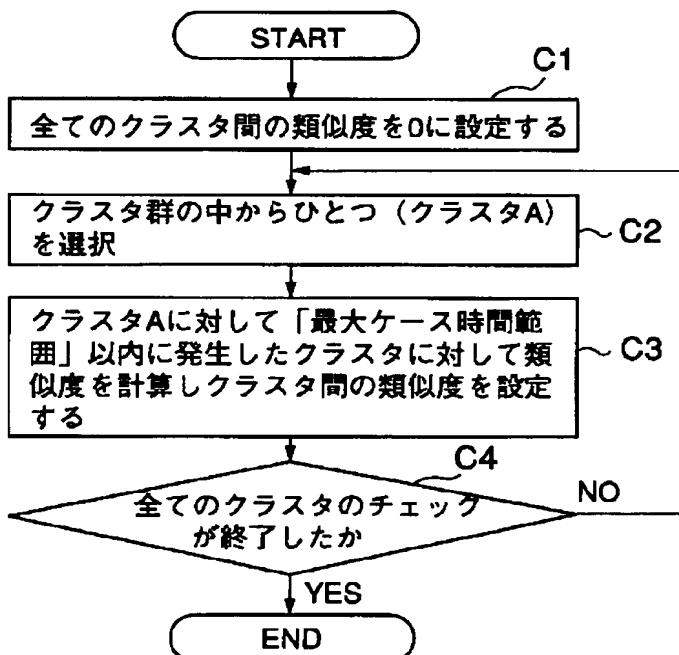
【図10】

クラスタ検索結果

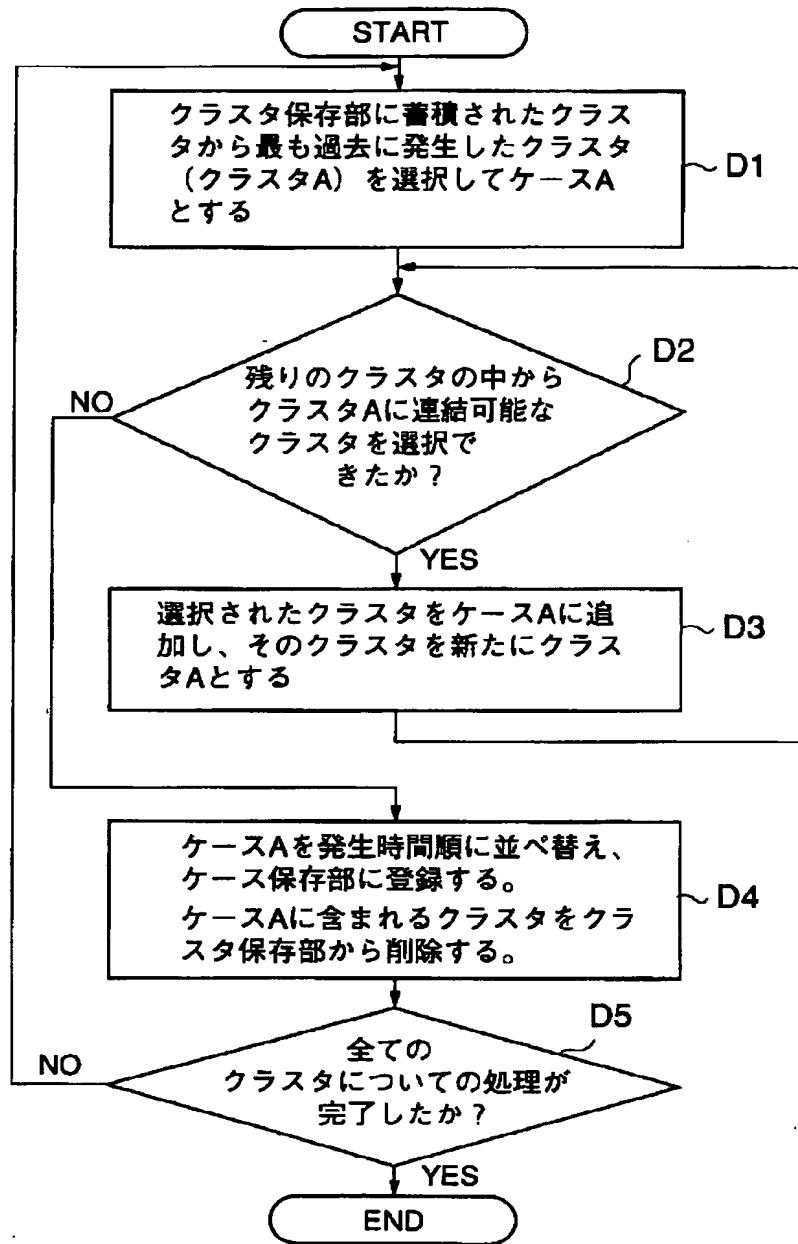
クラスタID	履歴ID	操作の種類	ドキュメント名	操作日時	ユーザ名
121	324	参照	C:\tex\ドキュメントA.doc	1997/12/24 10:30:20	鈴木
	330	印刷	C:\tex\ドキュメントA.doc	1997/12/24 10:43:45	鈴木
124	331	更新	F:\dirB\ドキュメントC.ppt	1997/12/24 10:50:45	鈴木

[閉じる]

【図11】



【図13】



フロントページの続き

(72)発明者 原口 竜也
神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

F ターム(参考) 5B075 ND03 ND20 NK04 NR10 NR12
PP02 PP03 PR03 PR04 PR06
UU06
5B082 DD04 FA11

This Page Blank (uspto)